Available online at www.sciencedirect.com

SCIENCE DIRECT®

GENOMICS

ACADEMIC PRESS

Genomics 82 (2003) 378–389

www.elsevier.com/locate/ygeno

High-throughput fingerprinting of bacterial artificial chromosomes using the SNaPshot labeling kit and sizing of restriction fragments by capillary electrophoresis

Ming-Cheng Luo,[a] Carolyn Thomas,[a] Frank M. You,[a] Joseph Hsiao,[b] Shu Ouyang,[b] C. Robin Buell,[b] Marc Malandro,[c] Patrick E. McGuire,[d] Olin D. Anderson,[e] and Jan Dvorak[a,]*

[a] *Department of Agronomy and Range Science, University of California at Davis, Davis, CA 95616, USA*
[b] *The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA*
[c] *Sagres Discovery, 2795 Second Street, Suite 400, Davis, CA 95616, USA*
[d] *Genetic Resources Conservation Program, University of California, Davis, CA 95616, USA*
[e] *USDA-ARS-WRRC-GGD, 800 Buchanan Street, Albany, CA 94710, USA*

Received 25 October 2002; accepted 27 April 2003

Abstract

We have developed an automated, high-throughput fingerprinting technique for large genomic DNA fragments suitable for the construction of physical maps of large genomes. In the technique described here, BAC DNA is isolated in a 96-well plate format and simultaneously digested with four 6-bp-recognizing restriction endonucleases that generate 3′ recessed ends and one 4-bp-recognizing restriction endonuclease that generates a blunt end. Each of the four recessed 3′ ends is labeled with a different fluorescent dye, and restriction fragments are sized on a capillary DNA analyzer. The resulting fingerprints are edited with a fingerprint-editing computer program and contigs are assembled with the FPC computer program. The technique was evaluated by repeated fingerprinting of several BACs included as controls in plates during routine fingerprinting of a BAC library and by reconstruction of contigs of rice BAC clones with known positions on rice chromosome 10.

Keywords: Fingerprint; Capillary electrophoresis; Contig; Physical mapping; BAC; Multiple digestion; High-throughput; End labeling

Arrayed libraries of large genomic DNA fragments are indispensable tools for genomic research. Of several large-insert cloning systems that have been developed, bacterial artificial chromosomes (BACs) [1] are the most versatile and currently are the most extensively used. The assets of the BAC cloning system are the relative ease with which BAC genomic libraries can be constructed, the stable maintenance of the large DNA inserts in the BAC vector, the easy arraying of libraries and screening, and the easy manipulation of clones using conventional molecular techniques.

The utility of arrayed BAC libraries would be greatly enhanced if clones would be arranged into contigs. A contig is a contiguous, gap-free order of overlapping clones reflecting the sequence of nucleotides in a chromosome [2,3]. Contigs can be physically superimposed on a chromosome to form a regional and global physical map of the chromosome.

Several strategies have been utilized for the construction of BAC contigs. One strategy, available to genome sequencing projects, is to sequence the ends of a large number of BAC clones, which is then followed by a global search for an identical sequence between the nucleotide sequence of a seed BAC and the BAC end sequences [4]. Contigs are built in a stepwise fashion as sequencing proceeds along a chro-

* Corresponding author. Fax: +1-530-752-4361.
E-mail address: jdvorak@ucdavis.edu (J. Dvorak).

0888-7543/03/$ – see front matter © 2003 Elsevier Science (USA). All rights reserved.
doi:10.1016/S0888-7543(03)00128-9

mosome. Another approach is to digest BAC clones with one or more restriction endonucleases and use the profiles of the resulting restriction fragments as fingerprints. A pair-wise search for overlaps between fingerprints is then used to build contigs [5]. Contigs of fingerprinted, large-insert clones can be constructed by targeting all or only a subset of restriction fragments in a search for overlaps. If all restriction fragments are targeted, large-insert clones are digested with a 6-bp-recognizing restriction endonuclease and the restriction fragments are sized on stained agarose gels [6]. The entire restriction profile of each clone is used during contig assembly. If only a limited subset of restriction fragments is targeted, clones are digested with a pair of restriction endonucleases consisting of a 6-bp-recognizing restriction endonuclease and 4-bp-recognizing restriction endonuclease, such as *Hin*dIII and *Sau*3A, respectively, and the *Hin*dIII sites are labeled to visualize the fragments [3]. The sizes of restriction fragments are determined by polyacrylamide gel electrophoresis and used for contig assembly [3]. The *Sau*3A endonuclease was replaced by the blunt-end generating *Hae*III endonuclease to facilitate digestion and labeling of restriction fragments in a single reaction [7]. Replacing the radioisotope labeling, which was the original restriction fragment visualization strategy [3], with fluorescent dye labeling allows for automation of the process as it is feasible to size restriction fragments using an automated DNA analyzer [8]. Multiplexing BAC digests produced independently with three different pairs of restriction endonucleases, each labeled with a different fluorescent dye, further improved the fluorescent dye fingerprinting method [9]. Alternatively, a single restriction endonuclease pair employing a type IIS restriction endonuclease can be employed. From 1 to 4 bases of the 5′ overhang are determined [10,11]. The information about the nucleotide sequence at the restriction site greatly reduces the likelihood of false-positive fragment sharing between the profiles of unrelated BACs [10].

We report here the development of a simple, high-throughput fingerprinting method for fully automated fingerprinting of BAC clones. The technique employs simultaneous restriction digestion of each BAC with one 4-bp and four 6-bp restriction endonucleases, labeling each of the 6-bp restriction sites with a different fluorescent dye in a single step, and sizing the fragments by capillary electrophoresis. The electronic files of the fingerprints are edited by an editing program that removes undesirable clones from the fingerprinted population and eliminates vector and background peaks from the profiles. Contigs are then assembled with the FPC computer program [5].

Results

Fingerprinting technique

Four 6-bp-recognizing restriction endonucleases, *Bam*HI, *Eco*RI, *Xba*I, and *Xho*I, each producing a 3′ recessed end, were used for BAC fingerprinting (Table 1). The 5′ overhang served as a template for the extension of the 3′ recessed end by Ampli*Taq* FS polymerase during the labeling reaction with the SNaPshot labeling kit (Applied Biosystems, Foster City, CA, USA; No. 4323155). The Ampli*Taq* FS extended the 3′ recessed end by adding an appropriate dideoxynucleotide labeled with a specific fluorescent dye. Since the four restriction sites differed in the first nucleotide on the 5′ overhang beyond the 3′ recessed end, the 3′ recessed end of each of the four restriction sites was labeled with a different fluorescent dye (Table 1). The fifth restriction endonuclease used was *Hae*III, which recognizes GGCC and produces blunt ends. The blunt ends of restriction fragments were not labeled by Ampli*Taq* FS. Fingerprinted BAC DNAs were ethanol precipitated, dissolved in formamide, and heat denatured, and LIZ 500 size standards (Applied Biosystems, No. 4322682, size range from 35 to 500 bp) were added. The sizes of labeled restriction fragments were determined by capillary electrophoresis (ABI 3100 Genetic Analyzer, Applied Biosystems). Data were collected for each 6-bp-recognizing restriction endonuclease in the 50 to 500 bp range.

Table 1
Characteristics of restriction sites and labeling of fragments

Restriction endonuclease	Restriction site	ddNTP	Fluorescent dye label	Color of restriction fragments
*Eco*RI	G ↓ AATTC	A	dR6G	Green
*Bam*HI	G ↓ GATCC	G	dR110	Blue
*Xba*I	T ↓ CTAGA	C	dTAMRA	Yellow
*Xho*I	C ↓ TCGAG	T	dROX	Red
*Hae*III	GG ↓ CC	None		

To digest each BAC DNA with the five restriction endonucleases simultaneously, the choice of the 6-bp-recognizing restriction endonucleases was guided by their buffer compatibility. The choice was also guided by the cost of enzymes. On the basis of these criteria, *Bam*HI, *Xba*I, and *Xho*I were selected for labeling of the 3′ recessed ends of restriction sites with ddGTP, ddCTP, and ddTTP, respectively (Table 1). Two restriction endonucleases, *Hin*dIII and *Eco*RI, were considered for labeling restriction fragments with ddATP. Four fully sequenced wheat BACs containing euchromatic DNA fragments were examined for the frequency of *Hin*dIII, *Eco*RI, *Bam*HI, *Xba*I, and *Xho*I restriction sites. It is widely known that *Hin*dIII cuts wheat DNA more often than *Eco*RI. There were more *Hin*dIII sites predicted by the nucleotide sequences than *Eco*RI sites in each of the four BACs. Across the four BACs, the *Hin*dIII sites outnumbered the *Eco*RI sites 2:1 (Table 2). Because too many fragments in a profile increase the frequency of false-positive matches between clones during contig assembly (see Contig assembly), *Eco*RI was chosen over *Hin*dIII for labeling 3′ recessed ends with ddATP. This decision was based on frequencies of restriction sites in euchromatic wheat BAC clones. The frequencies of restriction sites may

Table 2
Predicted numbers of restriction fragments in SNaPshot fingerprints of two *Triticum monococcum* and two *T. turgidum* BACs in the range of 50–500 bp

Restriction endonuclease	*T. monococcum* BACs		*T. turgidum* BACs		Total
	116F2	115G1	BAC1	BAC2	
*Eco*RI	31	38	32	32	133
*Bam*HI	20	36	53	32	141
*Xba*I	31	47	38	41	157
*Xho*I	26	30	46	23	125
Total	108	151	169	128	
*Hind*III	43	51	68	77	239

differ in other species or clones containing tandem repeated nucleotide sequences.

The digested and labeled BAC DNAs contained four classes of restriction fragments (Fig. 1). Most labeled restriction fragments were generated by DNA cleavage with a 6-bp-recognizing restriction endonuclease and *Hae*III (Class 1). Only a single strand was labeled in these fragments. The second class (Class 2) consisted of restriction fragments generated by DNA cleavage with the same 6-bp-recognizing restriction endonuclease. Both strands were labeled with the same fluorescent dye and were of identical lengths. The third class (Class 3) consisted of fragments produced by different 6-bp-recognizing restriction endonucleases. Both strands were of identical lengths but were labeled with different fluorescent dyes. The fourth class (Class 4) consisted of restriction fragments produced by *Hae*III cleavage alone. These were the most numerous fragments and were not labeled. Of the three classes labeled, only in Class 2 fragments was the relationship between the number of restriction fragments and the number of peaks in the electropherograms not 1 to 1. Peaks corresponding to Class 2 restriction fragments were examined in the profile of BAC 115G1. A total of four Class 2 fragments were predicted, accounting for 3.3% of all labeled restriction fragments, in BAC 115G1 by the nucleotide sequence of the clone. In a fingerprint randomly selected from a population of 326 fingerprints of this BAC (Fig. 2), in three of the four Class 2 fragments, the mobility of two labeled strands differed and generated two peaks. The fourth Class 2 fragment showed a single peak of an elevated height. An elevated height of a peak was by no means a diagnostic characteristic of Class 2 restriction fragments, since the heights of all peaks varied extensively (Fig. 2), presumably reflecting the variation with which Ampli*Taq* FS labels different restriction sites [8].

The sizes of restriction fragments observed in the profile of BAC 115G1 were compared with the sizes predicted from the nucleotide sequence of the clone (in parentheses in Fig. 2). The observed sizes of restriction fragments differed up to 7 bp from the predicted sizes (a maximum of 6 bp is seen in Fig. 2). The same observations were reported by other investigators who used fluorescent dyes for fingerprinting [9]. These deviations from expected mobility have no effect on the utility of the fingerprints for contig assembly because they are constant (see Reproducibility and tolerance). They also do not result in multiple peaks (Fig. 2). It would, nevertheless, be desirable to know the causes of these mobility shifts. It seems unlikely that the shifts were caused by secondary DNA structures, since DNA fragments were fractionated under denaturing conditions during capillary electrophoresis, which were expected to preclude the formation of secondary structures in single-stranded DNA. To assess if the mobility shifts were caused by differences in GC content among fragments, the GC contents of restriction fragments generated by digestion of BAC 115G1 were computed from the nucleotide sequence of the BAC clone. Only fragments for which the correspondence between a peak in the electropherogram and the predicted restriction fragment was unequivocal were used to assess correlation between fragment GC content and fragment mobility shift. The nonsignificant Pearson correlation coefficient $r = 0.06$

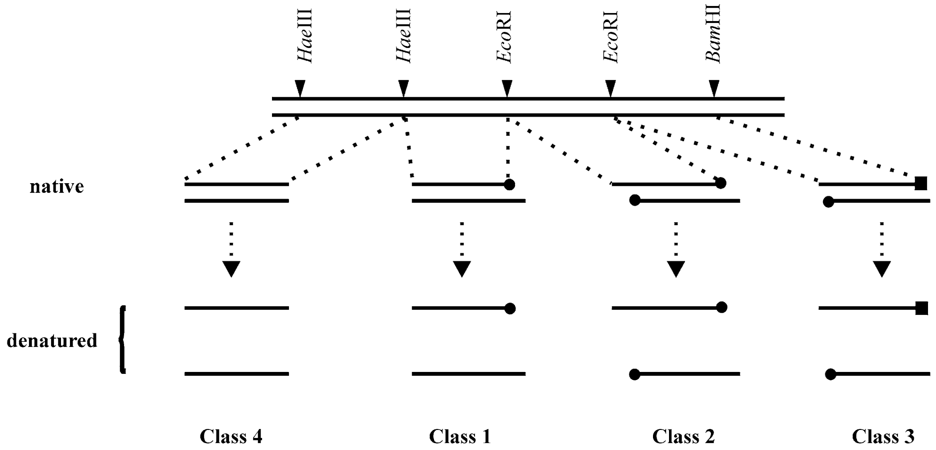

Fig. 1. A diagram of the origin of four classes of restriction fragments and their labeling with fluorescent dyes. The solid circles indicate labeled *Eco*RI sites and the solid square indicates the labeled *Bam*HI site.

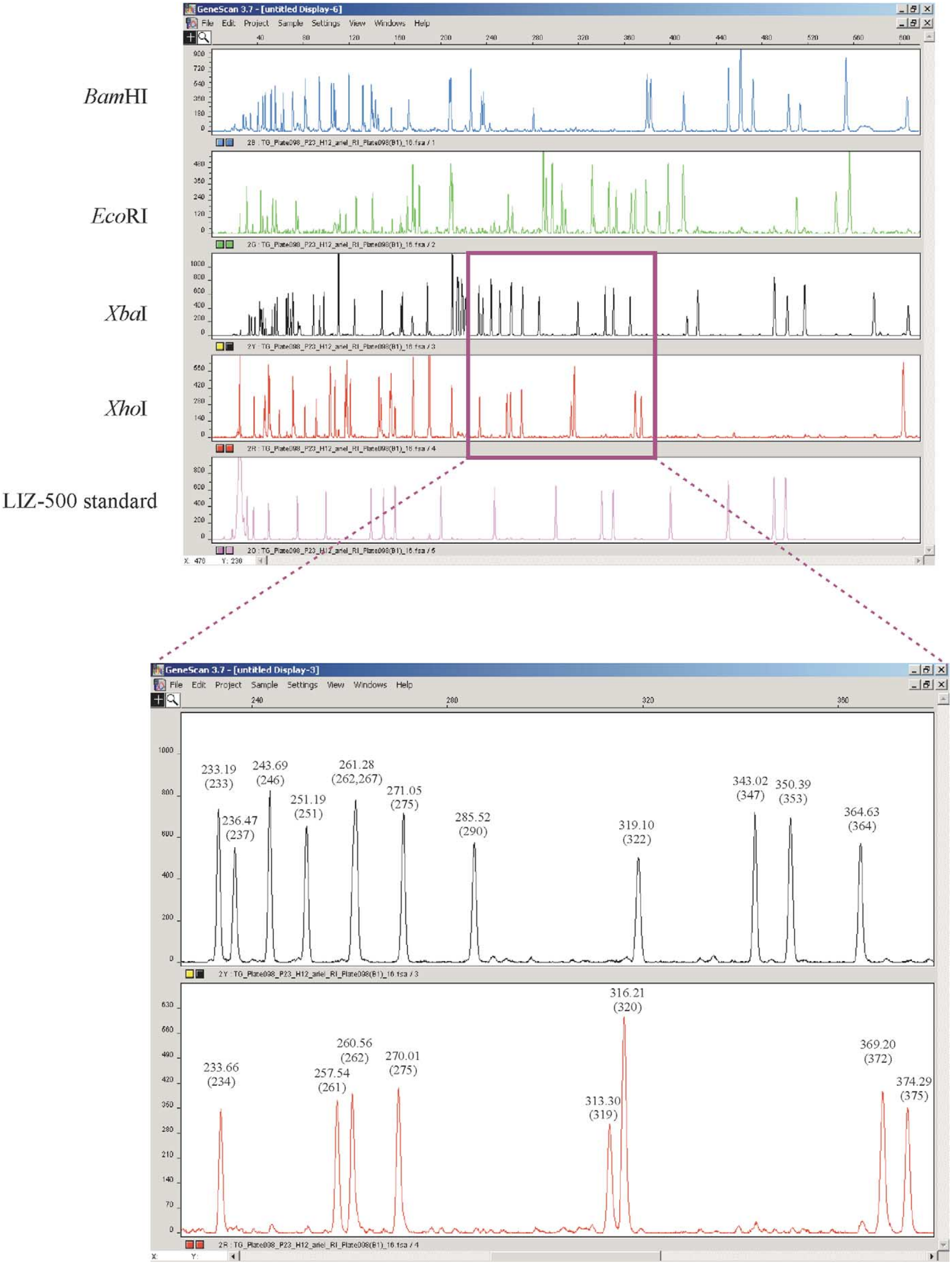

Fig. 2. *Bam*HI, *Eco*RI, *Xba*I, and *Xho*I electropherograms of *T. monococcum* BAC 115G1. The observed sizes of fragments and sizes expected from the nucleotide sequence (in parentheses) of the BAC are shown.

Table 3
Standard deviation (*S*) of fragment sizing in *N* replications of fingerprinting of BAC clones 115G1 and 116F2 and the sizes of confidence intervals

Clone	*N*	*S* (bp)	80% confidence interval (bp)	90% confidence interval (bp)	95% confidence interval (bp)
115G1	326	0.1442	$\bar{x} \pm 0.185$	$\bar{x} \pm 0.237$	$\bar{x} \pm 0.283$
116F2	342	0.1382	$\bar{x} \pm 0.177$	$\bar{x} \pm 0.227$	$\bar{x} \pm 0.271$

($p = 0.54$) indicated that mobility shifts of restriction fragments were not caused by variation in their GC content.

Methylation

Except for *Xba*I, the cleavage of BAC DNA by the remaining four restriction endonucleases is not affected by *Escherichia coli* DNA methylation. The 3′-most nucleotides (GA) of the *Xba*I restriction site are the first two nucleotides of the DAM methylation site (GATC), of which DAM methylase methylates the A. The likelihood that the A of an *Xba*I site will be methylated is therefore 1/16, assuming that the frequencies of T and C in BAC inserts is 1/4 each. Since *Xba*I is sensitive to A methylation, 1/16 of the *Xba*I sites will not be cleaved. The unavailability of another inexpensive 6-bp-recognizing restriction endonuclease with a G on the 5′ overhang adjacent to the recessed 3′ end led us to choose *Xba*I for fingerprinting, despite the sensitivity of this enzyme to *E. coli* methylation. To determine whether methylation of *Xba*I sites will introduce variation into the profiles, we tested a population of 326 fingerprints of BAC 115G1 for variation within the population. There were four methylation sites in this BAC. None of the 326 fingerprints showed any of the four predicted fragments. This provided empirical evidence that methylation of these rare *Xba*I sites is usually complete and will not introduce variation into the fingerprints.

Reproducibility and tolerance

To assess reproducibility of the fingerprinting technique during routine BAC fingerprinting, BACs 115G1 and 116F2 were included in 96-well plates of a BAC library being fingerprinted. DNA of the clones was isolated and fingerprinted over a period of 2 months, and fragments were sized using two ABI 3100 instruments over the same time period (Table 3). Fingerprinting of BAC 115G1 was repeated 326 times and that of BAC 116F2 342 times. Standard deviations (*S*) of fragment size estimates were highest in the 50 to 70 bp range and the 250 to 350 bp range (Fig. 3). Except for these two regions of the profiles, fragment size variation was independent of fragment length and was homogeneous as indicated by the absence of correlation between fragment size and *S* ($r = 0.07$, $p = 0.36$). The elevated standard deviations in the 250 to 350 bp range (Fig. 3) are probably caused by a large gap in the LIZ 500 standard created by the removal of the 250 bp marker from the profiles, which is advisable since the mobility of this fragment does not corresponds to its size under denaturing conditions [12]. It is expected that standard deviation of fragment sizing in the

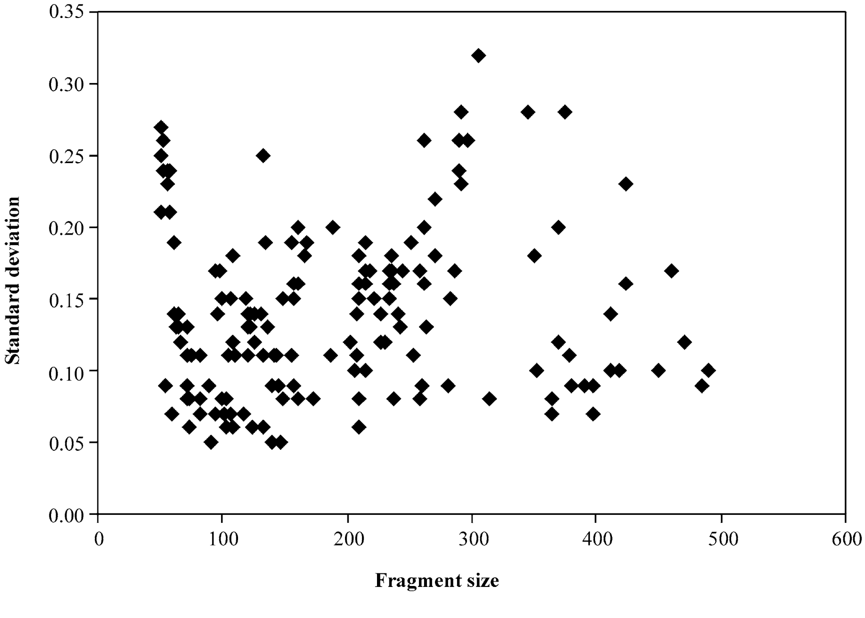

Fig. 3. Standard deviations in basepairs of restriction fragment size estimates in the range from 50 to 500 bp in the population of 326 fingerprints of BAC 115G1 and 342 fingerprints of BAC 116F2.

250 to 350 bp range will improve once a revised size standard is developed by Applied Biosystems. Because of large *S* in the 50- to 70-bp fragment size range, this region of profiles was not used in further analyses.

Mean standard deviation was computed for restriction fragments in the 70 to 500 bp range, and 80, 90, and 95% confidence intervals around the mean fragment size were computed (Table 3). An estimate of reproducibility of the routine fingerprinting with this technique can be obtained from these confidence intervals. For instance, the 90% confidence interval for the two clones indicates that if DNA is isolated independently twice from a BAC and fingerprinted, 90% of the fragments will be within 0.46 bp of one another across the entire 70 to 500 bp range and will be shared by the two profiles.

These predictions closely agree with empirical data on fragment sharing between different fingerprints of the same BAC clone at a specific tolerance level. Tolerance is defined as the size limits within which the observed sizes of two restriction fragments must be found to be considered identical during contig assembly by the FPC program [13]. The setting of tolerance is limited by variation in fragment sizing across time and instruments. By extrapolating from confidence intervals in Table 3, we find that a tolerance of 0.4 bp corresponds approximately to an 83% confidence interval. Hence, if a tolerance of 0.4 bp is used it is expected that 17% of restriction fragments will be outside the tolerance range and these fragments will not be considered the same by the FPC program. A fingerprint of BAC 116F2 is expected from its nucleotide sequence to have 108 restriction fragments and BAC 115G1 is expected to have 151 restriction fragments. Two fingerprints were randomly selected from the population of BAC 116F2 fingerprints and the number of shared restriction fragments in the 70 to 500 bp range was determined using a tolerance of 0.4 bp. The process was repeated 100 times each for BAC 116F2 and BAC 115G1 with replacements of fingerprints in the population. Two randomly chosen 116F2 BACs shared on average 79.3% restriction fragments and two 115G1 BACs shared on average 83.9% restriction fragments. These empirical estimates of restriction fragment sharing at 0.4 bp tolerance were close to the 83% predicted and suggested that the false-negative fragment sharing rate was 17% of restriction fragments. A restriction profile of an average BAC consists of approximately 120 restriction fragments (Table 2). If two average BACs overlap by, e.g., 50% of their restriction fragments, this false-negative rate predicts that only 50 of the 60 shared restriction fragments will actually be considered identical if a tolerance of 0.4 bp is used.

Contig assembly

The validity of contigs assembled from fingerprinted BAC clones depends on the fragment sizing accuracy, the reproducibility of the fingerprinting process, the information content of fingerprints, the extent of random matching of fragments during contig assembly (false-positive fragment sharing), and the presence of repeated nucleotide sequences. The effects of some of these variables on contig construction from SNaPshot fingerprinted BACs were assessed by fingerprinting rice BACs forming two chromosome 10 contigs. The contigs were originally constructed by the agarose fingerprinting method utilizing the *Hin*dIII restriction endonuclease [14] and assembled with FPC at the probability of coincidence (Sulston score) of 1×10^{-12}. The exact position of each clone in a contig was determined by BAC-end sequencing and superimposing the BAC ends on the finished sequence of chromosome 10. The overlap between BACs in the contig ranged from 15.4 to 100% of restriction fragments. Contig A contained 26 clones and contig B+C contained 58 clones. The two contigs were from different regions of rice chromosome 10.

Contigs were assembled using tolerances ranging from 0.4 to 0.6 bp and probabilities of coincidence ranging from 10^{-1} to 10^{-36}. Using a 0.4 bp tolerance, the 84 clones clustered into two contigs at Sulston scores ranging from 9×10^{-4} to 1×10^{-31}. At Sulston scores higher than 9×10^{-4}, the two contigs collapsed into a single contig. At Sulston scores below 1×10^{-31}, the B+C contig disintegrated into two contigs because of the failure to join BACs 23 and 24, which overlap by 15.4% of nucleotides. When tolerance was increased to 0.5 and 0.6 bp, two contigs were assembled at Sulston scores from 9×10^{-3} to 9×10^{-27} and 7×10^{-5} to 1×10^{-25}, respectively. The greater range of the probability of coincidence at which the two contigs were faithfully reconstructed suggested that tolerance of 0.4 is superior to tolerance of 0.5 or 0.6 for contig assembly.

The order of BACs fingerprinted by the SNaPshot fingerprinting method in contigs assembled at coincidence 1×10^{-25} and tolerance of 0.4 bp was compared with the order of BACs in contigs based on the nucleotide sequence of chromosome 10 (Figs. 4 and 5). The general order of clones fingerprinted with the SNaPshot fingerprinting method was similar to that in the chromosome 10 sequence-based contigs, although small inversions of BAC order were present within both contigs. Since the position of the clones in the contigs based on the nucleotide sequence of chromosome 10 was unequivocal, these small differences in the order of the BACs almost certainly reflected imperfect estimation in the overlap lengths between BAC clones in the contigs assembled with the SNaPshot fingerprinting method.

The correspondence of the BAC order in contigs relative to the chromosome 10 nucleotide sequence-based order was used to measure the effects of various parameters on the fidelity of contig assembly. To express numerically changes in the BAC order, the number of BACs by which a BAC was displaced relative to the chromosome 10 nucleotide sequence-based BAC order was counted and summed across all displaced BACs in a contig. This sum will be referred to as BAC displacement index of a contig. If the sequence of BACs in a contig were the same as the chro-

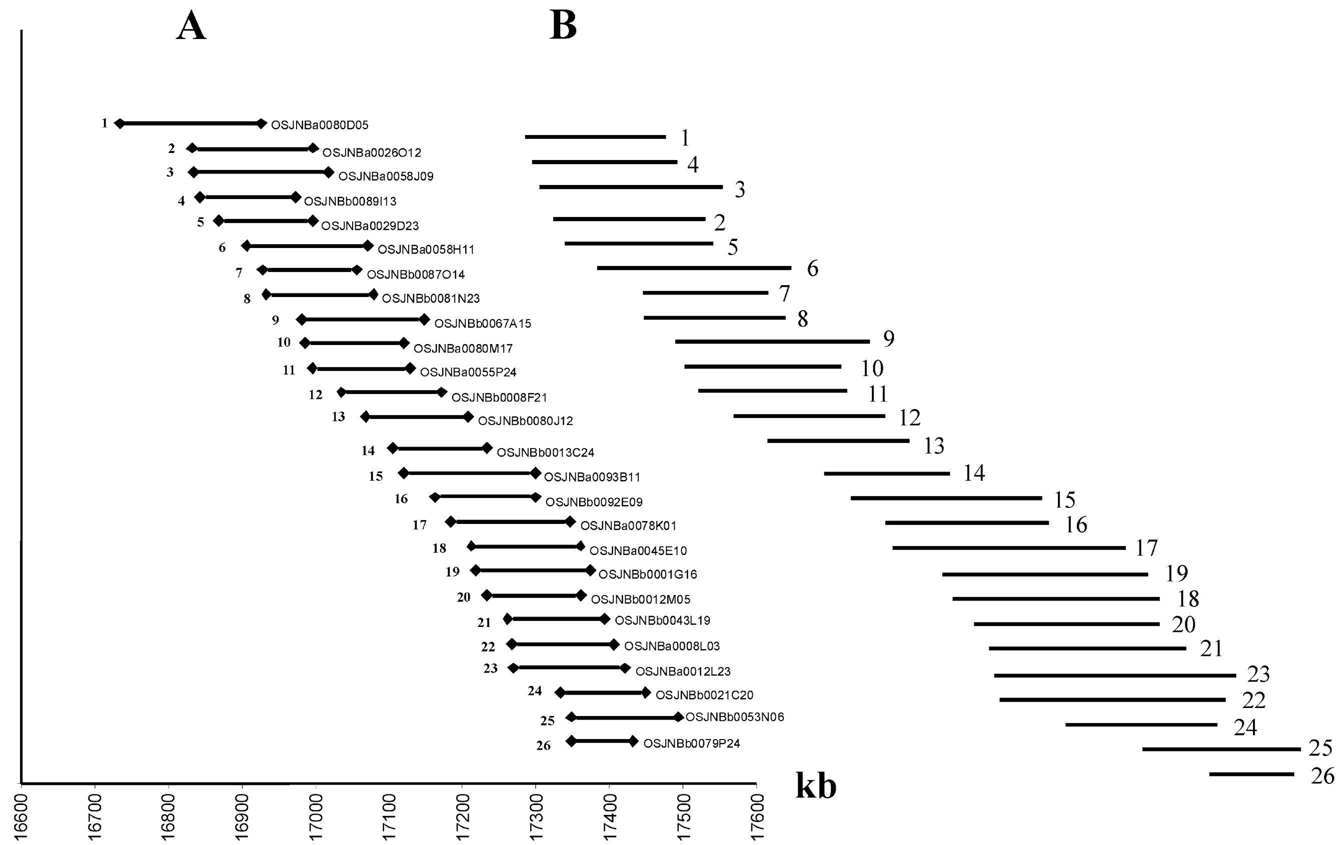

Fig. 4. Contig A of rice BAC clones from chromosome 10 based on the alignment of BAC-end nucleotide sequences to the complete, finished sequence of rice chromosome 10 (A) and SNaPshot fingerprinting method (B). BACs are oriented in the 5′–3′ direction. (A) The horizontal axis represents the nucleotide sequence of rice chromosome 10 (in kb). The relative positions of BACs in the contig were inferred from comparison of their end sequences with the nucleotide sequence of rice chromosome 10. The designations of individual BACs are to the right of the clones. Arbitrary designations reflecting the sequence of BACs in the contig based on nucleotide sequence are to the left of the clones. (B) The contig was assembled from a database of fingerprints of clones including contigs A and B+C (see Fig. 5). Tolerance of 0.4 bp and Sulston score 1×10^{-25} were used during contig assembly. Note that only clones of contig A are present and their order is similar to the order of clones in the contig shown in (A).

mosome 10 nucleotide sequence-based BAC order, the index would be zero. The larger the BAC displacement index, the less perfect the match between the assembled contig and chromosome 10 nucleotide sequence-based BAC order.

The effects of the amount of information gathered about BACs by fingerprinting on the fidelity of contigs were assessed by comparing BAC displacement indexes of contigs assembled using BACs fingerprinted with one, two, and four 6-bp restriction endonucleases. BAC displacement indexes of the four 6-bp restriction endonuclease, four-color SNaPshot fingerprinting technique were lower than the mean BAC displacement indexes of contigs assembled using only the single 6-bp restriction endonuclease-based technique or two 6-bp restriction endonuclease, two-color based technique (Table 4). BAC displacement indexes for the four 6-bp restriction endonuclease, four-color technique were 12 and 22 for the A and B+C contigs, respectively. In contrast, mean BAC displacement indexes for the single 6-bp restriction endonuclease-based technique were 18 and 105 for the A and B+C contigs, respectively. Mean BAC displacement indexes for two 6-bp restriction endonuclease, two-color-based technique were 13 and 47 for the A and

Fig. 5. (A) Contig B+C of rice BAC clones from chromosome 10 based on alignment of BAC-end nucleotide sequences to the complete, finished sequence of rice chromosome 10. The horizontal axis represents the nucleotide sequence of rice chromosome 10. The relative positions of BACs in the contig were inferred by comparison of their end sequences with the nucleotide sequence of rice chromosome 10. BACs are oriented in the 5′–3′ direction. The actual designations of individual BACs are to the right of the clones. Arbitrary designations reflecting the sequence of BACs in the contig are to the left of the clones. (B) Contig B+C of rice BAC clones from chromosome 10 constructed by the SNaPshot fingerprinting method. The contig is oriented in the 5′–3′ direction. BACs are designated by arbitrary designations reflecting their position in the contig shown in (A). The contig was assembled from a database of fingerprints of clones including contigs A (see Fig. 4) and B+C. Tolerance of 0.4 bp and Sulston score 1×10^{-25} were used during contig assembly. Note that only clones of contig B+C are present and their order is similar to the order of clones in the contig shown in (A). The Sulston score 9×10^{-32} in (A) indicates that contig B and contig C separate at that level.

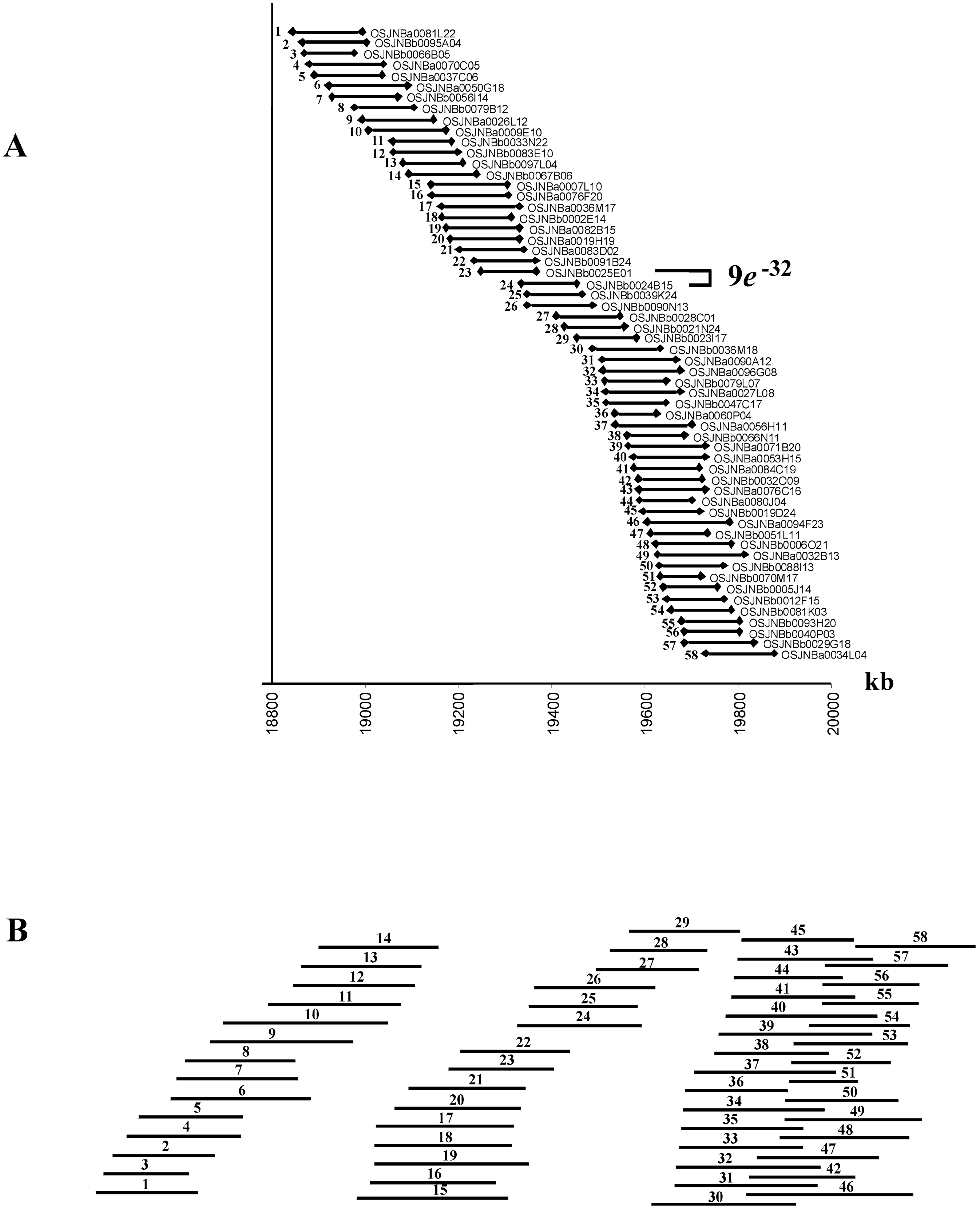
A
1 OSJNBa0081L22
2 OSJNBb0095A04
3 OSJNBb0066B05
4 OSJNBa0070C05
5 OSJNBa0037C06
6 OSJNBa0050G18
7 OSJNBb0056I14
8 OSJNBb0079B12
9 OSJNBa0026L12
10 OSJNBa0009E10
11 OSJNBb0033N22
12 OSJNBb0083E10
13 OSJNBb0097L04
14 OSJNBb0067B06
15 OSJNBa0007L10
16 OSJNBa0076F20
17 OSJNBa0036M17
18 OSJNBb0002E14
19 OSJNBa0082B15
20 OSJNBa0019H19
21 OSJNBa0083D02
22 OSJNBb0091B24
23 OSJNBb0025E01
24 OSJNBb0024B15
25 OSJNBb0039K24
26 OSJNBb0090N13
27 OSJNBb0028C01
28 OSJNBb0021N24
29 OSJNBb0023I17
30 OSJNBb0036M18
31 OSJNBa0090A12
32 OSJNBa0096G08
33 OSJNBb0079L07
34 OSJNBa0027L08
35 OSJNBb0047C17
36 OSJNBa0060P04
37 OSJNBa0056H11
38 OSJNBb0066N11
39 OSJNBa0071B20
40 OSJNBa0053H15
41 OSJNBa0084C19
42 OSJNBb0032O09
43 OSJNBa0076C16
44 OSJNBa0080J04
45 OSJNBb0019D24
46 OSJNBa0094F23
47 OSJNBb0051L11
48 OSJNBb0006O21
49 OSJNBa0032B13
50 OSJNBb0088I13
51 OSJNBb0070M17
52 OSJNBb0005J14
53 OSJNBb0012F15
54 OSJNBb0081K03
55 OSJNBb0093H20
56 OSJNBb0040P03
57 OSJNBb0029G18
58 OSJNBa0034L04
9e-32
kb
18800
19000
19200
19400
19600
19800
20000
B
14
13
12
11
10
9
8
7
6
5
4
2
3
1
22
23
21
20
17
18
19
16
15
29
28
27
26
25
24
45
43
44
41
40
39
38
37
36
34
35
33
32
31
30
58
57
56
55
54
53
52
51
50
49
48
47
42
46

Table 4
BAC displacement indexes in contig A and contig B+C fingerprinted with single 6-bp restriction endonuclease, single-color technique (*Bam*HI, *Eco*RI, *Xba*I, and *Xho*I); two 6-bp restriction endonuclease, two-color technique (*Bam*HI + *Eco*RI and *Xba*I + *Xho*I); and four 6-bp restriction endonuclease, four-color technique (*Bam*HI + *Eco*RI + *Xba*I + *Xho*I)

Restriction profile	Contig A	Contig B+C
*Bam*HI	12	134
*Eco*RI	18	124
*Xba*I	8	52
*Xho*I	34	110
*Bam*HI + *Eco*RI	12	54
*Xba*I + *Xho*I	14	40
*Bam*HI + *Eco*RI + *Xba*I + *Xho*I	12	22

B+C contigs, respectively. Hence, BAC displacement indexes were the lowest for contigs constructed from BAC fingerprints produced with the four 6-bp restriction endonuclease, four-color technique.

There were marked differences among the BAC displacement indexes of the four single 6-bp restriction endonuclease-based contigs (Table 4). BAC displacement indexes were the lowest for *Xba*I in both A and B+C contigs. Since only two contigs were investigated, it is not clear to what extent this pattern was due to chance and to what extent it was a general property of the *Xba*I restriction endonuclease.

There was a good agreement between the sizes of clone overlaps measured by the number of shared restriction fragments and by alignment of the BAC end sequence against the chromosome 10 nucleotide sequence (Fig. 6). The relationship between these variables was approximately linear in the range from zero to about 80% overlap and then it plateaued (Fig. 6). The corollary of this relationship is that the number of restriction fragments in the fingerprint approximates the length of a clone or contig. Using the four wheat BAC clones and 84 rice BAC clones, this relationship was estimated to be 1.23 kb/fragment in wheat and 1.38 kb/fragment in rice.

Mean overlap between 178 randomly chosen pairs of unrelated BACs (those showing a zero overlap at the nucleotide sequence level) was 2.71 ± 0.18% and ranged from 0.49 to 5.85% of the fragments. Note that vector-derived fragments were removed from the profiles by the editing program. Assuming that there was no internal redundancy within and between the A and B+C rice contigs, these numbers estimated the false-positive fragment sharing rate of the four 6-bp restriction endonuclease, four-color technique.

Increase in the number of restriction fragments in a profile is expected to result in increases in the false-positive fragment sharing rate. To test this hypothesis, two fingerprints (*Bam*HI and *Eco*RI) produced by a single restriction endonuclease were combined into a single profile, generating a virtual two 6-bp restriction endonuclease, single-color fingerprint. Two randomly selected BACs, one from contig A and the other from contig B+C, were paired and the number of matched fragments were determined. The process was repeated 202 times. While a mean of 4.53 ± 0.29 restriction fragments (5.27%), ranging from 0.0 to 17.4% of the fragments, were matched by chance with the two 6-bp restriction endonuclease, single-color method, only 2.41 ± 0.21 restriction fragments (2.73%), ranging from 0.0 to 8.0% of the fragments, were matched by chance with the two 6 bp-restriction endonuclease, two-color method. The number of restriction fragments in the profiles was the same, but the length of the latter profiles was double the length of the former profiles. The increased density of restriction fragments in the profiles of the two 6-bp restriction endonuclease, single-color technique compared to the two 6-bp restriction endonuclease, two-color technique significantly increased the false-positive fragment matching rate (t = 25.8, $p < 0.001$).

Discussion

During routine fingerprinting spanning 2 months and performed on two ABI 3100 capillary DNA sequencers, the standard deviation of the fragment sizing suggested that 0.4, 0.5, and 0.6 bp tolerances could be used during contig assembly. Using the 0.4 bp tolerance level, two rice BAC contigs were faithfully assembled at Sulston scores ranging from 9×10^{-4} to 1×10^{-31}. Increasing the tolerance to 0.5

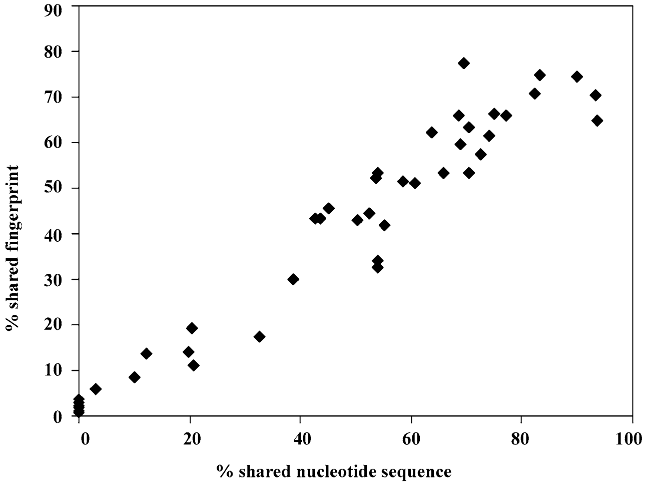

Fig. 6. Relationship between sharing of nucleotide sequence between randomly selected rice BACs in contigs A and B+C and sharing of the fingerprint generated by the SNaPshot fingerprinting method between the same rice BACs. The percentage of the sharing was computed with the formula $[S/(A + B - S)] \times 100$, where A represents the length of one BAC clone in terms of nucleotide sequence or the number of fragments in terms of fingerprints, B represents the length of the second BAC clone in terms of nucleotide sequence or the number of fragments in terms of fingerprints, S represents the length of shared nucleotide sequence or number of shared fragments with tolerance of 0.4 bp by the pair. Note that six pairs of randomly selected unrelated BACs share from 1 to 4% of their fingerprints.

and 0.6 bp impacted negatively this range. Therefore, it is preferable to use the 0.4 bp tolerance if fragments are sized by capillary electrophoresis compared to the 0.5 bp tolerance recommended earlier for gel-based electrophoresis sizing [9,11]. High fragment sizing error rate in the 50 to 70 bp range suggested that this fragment size range should be avoided to keep the overall fragment sizing error rate low.

The fidelity of contigs assembled from fingerprinted BACs depends, among other factors, on the rates of false-positive and false-negative restriction fragment matching. The false-negative matching rate, i.e., the failure to detect an existing match, depends on the sizing error rate and the choice of tolerance level. The larger the tolerance, the smaller the false-negative matching rate. The false-negative matching rate was about 17% using 0.4 bp tolerance when restriction fragments were sized by ABI 3100 capillary DNA analyzers. The choice of tolerance level has the opposite effect on the false-positive matching rate. The larger the tolerance, the larger the false-positive matching rate. Using 0.4 bp tolerance, the average empirical false-positive matching rate was 2.71% restriction fragments with the four 6-bp restriction endonuclease, four-color fingerprinting technique.

If the sizing range remains constant (e.g., 70 to 500 bp), the false-positive matching rate per restriction fragment increases with the increase in the number of labeled restriction fragments in the profile. Doubling the number of restriction fragments in the 70 to 500 bp range resulted in doubling the false-positive matching rate per restriction fragment. For this reason, we elected to use *Eco*RI restriction endonuclease instead of *Hin*dIII for labeling of the A, since *Hin*dIII produced up to twice as many fragments in the profile as *Eco*RI. For this same reason, a two 6-bp restriction endonuclease, single-color BAC fingerprinting technique [15] is expected to suffer from increased frequency of random matches, while simultaneously providing less information about each BAC, which is expected to reduce the fidelity of contigs. The fidelity of contigs assembled from BACs fingerprinted with the four 6-bp restriction endonuclease, four-color BAC fingerprinting technique described here would also be likely higher than the fidelity of contigs assembled from BACs labeled with a three 6-bp restriction endonuclease, three-color BAC fingerprinting technique reported by Ding et al. [9]. The technique developed here cannot be easily compared with that utilizing a type IIS restriction endonuclease [10,11]. That technique uses a single restriction enzyme and produces on average 36 fragments per BAC, whereas the present technique produces on average 120 fragments per BAC. This lower information level provided by the type IIS restriction endonuclease technique is expected to affect negatively the fidelity of contigs, as was demonstrated here. It is uncertain to what extent this disadvantage is offset by the determination of the sequence of four nucleotides adjacent to the cleavage point by a type IIS endonuclease, which minimizes the likelihood of false-positive matches.

The Sulston score determines the minimum relative size of the overlap needed for contiging two clones; the lower the Sulston score the larger must be the overlap. The undesirable consequence of using low Sulston scores is that more clones must be fingerprinted to obtain sizable contigs. The observation that the rice contigs were correctly assembled in a wide range of scores illustrates the sensitivity of the technique and its ability to detect correctly overlaps between BACs under a range of conditions. This flexibility is important since biological factors, such as genome size and redundancy caused by repeated nucleotide sequences, ultimately determine the minimum overlap level and hence Sulston scores that can be used during contig assembly. To obtain adequate coverage of a large genome, a large number of clones must be fingerprinted and utilized during contig assembly. A large number of pairwise comparisons during contig assembly make false-positive matches due to chance more relevant. Low tolerance values and low Sulston score values, i.e., long overlaps, must be used to counteract false-positive contig assembly in large genomes due to these statistical and biological reasons. The fact that the present technique contiged the B and C regions of the B+C contig, which is held by a short overlap, all the way to 10^{-31} is important for the utility of this technique for the construction of BAC contigs of large genomes.

In most grasses, the intergenic space is filled with retroelements and other repeated nucleotide sequences, which typically are less than 10 kb long. Since these elements are usually nested in each other [16], they create a patchwork in individual BAC clones. While the likelihood of a small overlap between unrelated BACs is high, the individuality of the patchwork among BACs is likely to reduce the likelihood of false-positive matches if the overlaps are long. Obviously, imaginative strategies must be used during the construction of physical maps of large genomes to allow for using reasonably high Sulston scores, thereby obviating the need for fingerprinting excessive numbers of BAC clones. One possibility is to search the entire database of fingerprints for restriction fragments originating from major classes of repeated nucleotide sequences. These are marked by significantly elevated frequencies in the database compared to the average fragment frequency across a database [17]. These restriction fragments can be eliminated from the fingerprint database during contig assembly. The program we developed for automated BAC fingerprint editing can be used to remove such fragments from fingerprints prior to contig assembly with FPC.

The present technique is well suited to fully automated fingerprinting procedures. By performing the digestion with five restriction endonucleases in a single step, the technique simplified the fingerprinting reactions relative to the multiplexing technique reported by Ding et al. [9]. The use of robots for BAC DNA isolation and fingerprinting and automated fragment sizing by capillary electrophoresis DNA analyzers makes it feasible to process a thousand or more BAC clones per day. For example, two technicians fingerprinted 170,000 BACs in 9 months, which is an annual throughput of 227,000 BACs, using the Qiagen R.E.A.L 96-Prep kit (Valencia, CA, USA) for BAC DNA isolation, the Tecan Genesis 150 robot for performing restriction digestion and labeling reactions in the 96-well format, and two 16-capillary ABI 3100 genetic analyzers (M.-C. Luo and J. Dvorak, unpublished). Utilization of a robot for BAC DNA isolation, such as Autogen 960, and replacement of ABI 3100 DNA analyzers with 48- or 96-capillary instruments, such as ABI 3730 or ABI 3730X, would increase the annual (250 working days) throughput of two workers to 500,000 BAC clones. The development of a software package (GenoProfiler, available at http://wheat.pw.usda.gov/PhysicalMapping/) for automated fingerprint editing minimizes the time and labor needed for manipulation of fingerprinting files upstream of FPC. The current cost of a fingerprint, including supplies and labor for BAC DNA isolation and fingerprinting, is ca. $3.50 per BAC. Considering these throughput levels and the costs, it seems that this technique opens a door for routine contiging of clones in BAC libraries of most organisms, including those with large genomes, such as wheat.

Materials and methods

BAC clones

Two fully sequenced BAC clones (116F2 and 115G1) of *Triticum monococcum* and two fully sequenced BAC clones of *Triticum turgidum* (BAC1 and BAC2) were provided by J. Dubcovsky (University of California, Davis, CA, USA). Clones 116F2, 115G1, BAC1, and BAC2 were 107.3, 128.6, 173.4, and 147.6 kb in length, respectively. Clones 116F2 and 115G1 overlap by 20.6 kb, whereas BAC1 and BAC2 overlap by 29.7 kb. A set of 84 rice (*Oryza sativa* spp. *japonica*) cv. Nipponbare BAC clones (average insert size 141.5 kb) was used to validate the fingerprinting procedure. These clones formed two different contigs. Contig A, generated by *Hin*dIII fingerprinting (Clemson University Genomic Institute (CUGI) [14]), consisted of 26 BAC clones that have been anchored to rice chromosome 10 at 48.4 cM (C.R. Buell, unpublished information). The remaining 58 BAC clones are within CUGI contigs B and C and are anchored on rice chromosome 10 at 58.6 cM. Contigs B and C overlap by two BAC clones, forming a single large contig termed B+C. BAC end sequences were used to align all 84 BAC clones on the genomic sequence of rice chromosome 10. Contig A represents 0.759 Mb of unique sequence, whereas contig B+C represents 1.033 Mb of unique sequence.

Fingerprinting reaction

A 96-well block containing 1.2 ml of 2× YT medium [18] with 12.5 μg/ml chloramphenicol per well was inoculated with a 96-pin replicator. The cultures were grown for 24 h on a HiGro shaker (Gene Machines, Inc., San Carlos, CA, USA) at 425 rpm, 37°C. BAC DNA was isolated with the Qiagen R.E.A.L 96-Prep kit either manually or by Qiagen robot following a procedure recommended by the manufacturer. Typically, 0.5–1.2 μg of DNA was obtained per BAC clone.

BAC DNA was dissolved in 42 μl of double-distilled (dd) water at 4°C overnight. A total of 9.0 μl of a solution containing 5 units each of *Bam*HI, *Eco*RI, *Xba*I, *Xho*I, and *Hae*III restriction endonucleases; 1× NEBuffer 2; 5 μg bovine serum albumin; 5 μg DNase-free RNase A (Sigma R-6513); and 0.1% β-mercaptoethanol was added and the DNA was digested at 37°C for 3 h. The digested DNA was transferred into a 96-well PCR plate compatible with the ABI 3100 Genetic Analyzer (Applied Biosystems). Each well contained 10.0 μl of the SNaPshot labeling solution (1 μl of SNaPshot Multiplex Ready Reaction Mix (ABI), 2 μl NEBuffer 2, 2.5 μl 100 mM Tris, pH 9.0, and 4.5 μl ddH_2O). The plates were briefly centrifuged at 100*g* and incubated at 65°C for 60 min.

A total of 6.0 μl of 2.5 M sodium acetate (pH 5.2) and 100 μl of prechilled 95% ethanol (−20°C) was added and plates were kept at −80°C for 10–15 minutes. DNA was sedimented in the plates at 3650*g* for 30 min, washed with 70% ethanol, and sedimented again at 2500*g* for 10 min. The plates were then turned upside down on a paper towel and centrifuged in that position at 500 rpm for 2 min. The sedimented DNA was air dried for 5 min.

Capillary electrophoresis

Dried DNA was dissolved in 10 μl of Hi-Di formamide (ABI No. 4311320), and 0.2 μl of ABI internal size standard LIZ-500 (ABI No. 4322682, size range from 35 to 500 bp) was added into each sample. The DNA in the plates was denatured at 95°C for 5 min and placed immediately on ice until loaded into an ABI 3100 DNA sequencer. Capillary electrophoresis was performed with 36-cm capillaries using the ABI default GeneScan module (ABI, 3100 POP-4, ABI Buffer No. 402824).

FPC data processing

Peak areas, peak heights, and fragment sizes in each BAC fingerprint profile were collected by the ABI Data Collection program without the 250-bp fragment in the size-standard file. The data off the ABI 3100 Genetic Analyzer were processed by the computer software package GenoProfiler (available at http://wheat.pw.usda.gov/PhysicalMapping/). This software package was used to distinguish peaks corresponding to restriction fragments from peaks generated by background noise in the profile of each BAC fingerprint and to remove vector restriction fragments from the profiles. The program also removed substandard profiles that may negatively affect contig assembly. The files generated by GenoProfiler were used in the FPC contig assembly.

Acknowledgments

This publication is based upon work supported by National Science Foundation Grant DBI-0077766. Funding for the work on rice chromosome 10 at TIGR was provided by a grant to C.R.B. from the U.S. Department of Agriculture (99-35317-8275), the National Science Foundation (DBI-998282), and the U.S. Department of Energy (DE-FG02-99ER20357). The authors express their thanks to Michele Morgante (DuPont) for many valuable suggestions.

References

[1] H. Shizuya, et al., Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector, Proc. Natl. Acad. Sci. USA 89 (1992) 8794–8797.

[2] R. Staden, A new computer method for the storage and manipulation of DNA gel reading data, Nucleic Acids Res. 8 (1980) 3673–3694.

[3] A. Coulson, J.S. Sulston, S. Brenner, J. Karn, Toward a physical map of the genome of the nematode *Caenorhabditis elegans*, Proc. Natl. Acad. Sci. USA 83 (1986) 7821–7825.

[4] J.C. Venter, H.O. Smith, L. Hood, A new strategy for genome sequencing, Nature 381 (1996) 364–366.

[5] C. Soderlund, S. Humphray, A. Dunham, L. French, Contigs built with fingerprints, markers, and FPCV4.7, Genome Res. 10 (2000) 1772–1787.

[6] M.V. Olson, et al., Random clone strategy for genomic restriction mapping in yeast, Proc. Natl. Acad. Sci. USA 83 (1986) 7826–7830.

[7] P.E. Klein, et al., A high-throughput AFLP-based method for constructing integrated genetic and physical maps: Progress toward a sorghum genome map, Genome Res. 10 (2000) 789–807.

[8] S.G. Gregory, G.R. Howell, D.R. Bentley, Genome mapping by fluorescent fingerprinting, Genome Res. 7 (1997) 1162–1168.

[9] Y. Ding, et al., Contig assembly of bacterial artificial chromosome clones through multiplexed fluorescence-labeled fingerprinting, Genomics 56 (1999) 237–246.

[10] S. Brenner, K.J. Livak, DNA fingerprinting by sampled sequencing, Proc. Natl. Acad. Sci. USA 86 (1989) 8902–8906.

[11] Y. Ding, et al., Five-color-based high-information-content fingerprinting of bacterial artificial chromosome clones using type IIS restriction endonucleases, Genomics 74 (2001) 142–154.

[12] Applied Biosystems, GeneScan Reference Guide. (2000) 5-6.

[13] C. Soderlund, I. Longden, R. Mott, FPC: a system for building contigs from restriction fingerprinted clones, CABIOS 13 (1997) 523–535.

[14] M.S. Chen, et al., An integrated physical and genetic map of the rice genome, Plant Cell 14 (2002) 537–545.

[15] Z. Xu, et al., An automated procedure for whole-genome physical mapping from large-insert BACs and BIBACs. Plant and Animal Genome XI Conference. (2003) 103 [Abstract].

[16] P. Sanmiguel, et al., Nested retrotransposons in the intergenic regions of the maize genome, Science 274 (1996) 765–768.

[17] M.A. Marra, et al., High throughput fingerprint analysis of large-insert clones, Genome Res. 7 (1997) 1072–1084.

[18] J. Sambrook, E.F. Fritsch, T. Maniatis, Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Press, Cold Spring Harbor, NY, 1989.